

STEIN'S METHOD AND THE LAPLACE DISTRIBUTION

John Pike

Department of Mathematics
University of Southern California

jpike@usc.edu

Haining Ren

Department of Mathematics
University of Southern California

hren@usc.edu

ABSTRACT. Using Stein's method techniques, we develop a framework which allows one to bound the error terms arising from approximation by the Laplace distribution and apply it to the study of random sums of mean zero random variables. As a corollary, we deduce a Berry-Esseen type theorem for the convergence of certain geometric sums. Our results make use of a second order characterizing equation and a distributional transformation which can be realized by iterating the $X - P$ bias transformation introduced by Goldstein and Reinert.

1. BACKGROUND AND INTRODUCTION

Beginning with the publication of Charles Stein's seminal paper [16] in 1972, probabilists and statisticians have developed a wide range of techniques based on characterizing equations for bounding the distance between the distribution of a random variable X and that of a random variable Z having some specified target distribution. The metrics for which these techniques are applicable are of the form $d_{\mathcal{H}}(\mathcal{L}(X), \mathcal{L}(Z)) = \sup_{h \in \mathcal{H}} |E[h(X)] - E[h(Z)]|$ for some suitable class of functions \mathcal{H} , and include as special cases the Wasserstein, Kolmogorov, and total variation distances. (The Kolmogorov distance gives the L^∞ distance between the associated distribution functions, so $\mathcal{H} = \{1_{(-\infty, a]}(x) : a \in \mathbb{R}\}$. The total variation and Wasserstein distances correspond to letting \mathcal{H} consist of indicators of Borel sets and 1-Lipschitz functions, respectively.) The basic idea is to find an operator \mathcal{A} such that $E[(\mathcal{A}f)(X)] = 0$ for all f belonging to some sufficiently large class of functions \mathcal{F} if and only if $\mathcal{L}(X) = \mathcal{L}(Z)$. For example, Stein showed that $E[(\mathcal{A}_N f)(Z)] = E[Zf(Z) - \sigma^2 f'(Z)] = 0$ for all absolutely continuous functions f such that these expectations exist if and only if $Z \sim N(0, \sigma^2)$ [16], and his student Louis Chen showed shortly thereafter that $Z \sim \text{Poisson}(\lambda)$ if and only if $E[(\mathcal{A}_P f)(Z)] = E[Zf(Z) - \lambda f(Z+1)] = 0$ for all functions f for which the expectations exist [2]. Similar characterizing operators have since been worked out for several other distributions [6, 11, 1, 13, 15]. Given such an operator \mathcal{A} for $\mathcal{L}(Z)$, one can consider the solution $f_h \in \mathcal{F}$ to the equation $(\mathcal{A}f)(x) = h(x) - E[h(Z)]$ for $h \in \mathcal{H}$. Taking expectations, absolute values, and suprema gives

$$d_{\mathcal{H}}(\mathcal{L}(X), \mathcal{L}(Z)) = \sup_{h \in \mathcal{H}} |E[h(X)] - E[h(Z)]| = \sup_{h \in \mathcal{H}} |E[(\mathcal{A}f_h)(X)]|.$$

The intuition is that since $E[(\mathcal{A}f)(Z)] = 0$ for $f \in \mathcal{F}$, the distribution of X should be close to that of Z when $E[(\mathcal{A}f)(X)]$ is close to zero. Remarkably, it is often easier to work with the right-hand side of the above

2010 *Mathematics Subject Classification.* 60F05.

Key words and phrases. Stein's Method, Laplace distribution, geometric summation.

equation, and the tools for analyzing distances between distributions in this manner are collectively known as Stein’s method. For more on this rich and fascinating subject, the authors recommend [17, 15, 3, 4].

In this paper, we apply the above ideas to the (symmetric) Laplace distribution. For $a \in \mathbb{R}$, $b > 0$, a random variable $W \sim \text{Laplace}(a, b)$ has cumulative distribution function

$$F_W(w; a, b) = \begin{cases} \frac{1}{2}e^{\frac{w-a}{b}}, & w \leq a \\ 1 - \frac{1}{2}e^{-\frac{w-a}{b}}, & w \geq a \end{cases}$$

and density

$$f_W(w; a, b) = \frac{1}{2b}e^{-\frac{|w-a|}{b}}, \quad w \in \mathbb{R}.$$

If $W \sim \text{Laplace}(0, b)$, then its moments are given by

$$E[W^k] = \begin{cases} 0, & k \text{ is odd} \\ b^k k!, & k \text{ is even} \end{cases}$$

and its characteristic function is

$$\varphi_W(t) = \frac{1}{1 + b^2 t^2}.$$

This distribution was introduced by P.S. Laplace in 1774, four years prior to his proposal of the “second law of errors,” now known as the normal distribution. Though nowhere near as ubiquitous as its younger sibling, the Laplace distribution appears in numerous applications, including image and speech compression, options pricing, and modeling sizes of sand particles, diamonds, and beans. For more properties and applications of the Laplace distribution, the reader is referred to the text [9].

Our interest in the Laplace distribution lies primarily in the fact that if X_1, X_2, \dots is a sequence of random variables (satisfying certain assumptions) and $N_p \sim \text{Geometric}(p)$ is independent of the X_i ’s, then the sum $p^{\frac{1}{2}} \sum_{i=1}^{N_p} X_i$ converges weakly to the Laplace distribution as $p \rightarrow 0$ [9]. Such geometric sums arise in a variety of settings [8], and the general setup (distributional convergence of sums of random variables) is exactly the type of problem for which one expects Stein’s method computations to yield useful results. Indeed, Erol Peköz and Adrian Röllin have applied Stein’s method arguments to generalize a theorem due to Rényi concerning the convergence of sums of a random number of positive random variables to the exponential distribution [12]. By an analogous line of reasoning, we are able to carry out a similar program for convergence of random sums of certain mean zero random variables to the Laplace distribution.

We begin in section 2 by introducing a Stein equation which we show completely characterizes the mean zero Laplace distribution. In section 3, we use this characterization to construct a distributional transformation which has the $\text{Laplace}(0, b)$ law as a fixed point, show how this transformation can be used to bound the distance to the Laplace distribution, and establish sufficient conditions for a distribution to lie in the domain of this transformation. In section 4, we apply these tools to the study of random sums and deduce a Berry-Esseen type result for the convergence of certain geometric sums to the Laplace distribution.

2. A CHARACTERIZING EQUATION

The first order of business is to find a characterizing equation for the Laplace distribution and establish some properties of its solutions. We begin with the following lemma.

Lemma 1. *Suppose that $W \sim \text{Laplace}(0, b)$, $b > 0$. Then if g is an absolutely continuous function such that g' is also absolutely continuous and $E[g(W)]$ exists, we have*

$$E[g(W)] - g(0) = b^2 E[g''(W)].$$

Proof. Applying Fubini's theorem twice shows that

$$\begin{aligned}
\int_0^\infty g''(x) e^{-\frac{x}{b}} dx &= \int_0^\infty g''(x) \left(\int_x^\infty \frac{1}{b} e^{-\frac{y}{b}} dy \right) dx = \frac{1}{b} \int_0^\infty \int_0^y g''(x) e^{-\frac{y}{b}} dx dy \\
&= \frac{1}{b} \int_0^\infty (g'(y) - g'(0)) e^{-\frac{y}{b}} dy = \frac{1}{b} \int_0^\infty g'(y) e^{-\frac{y}{b}} dy - \frac{g'(0)}{b} \int_0^\infty e^{-\frac{y}{b}} dy \\
&= \frac{1}{b} \int_0^\infty g'(y) \left(\int_y^\infty \frac{1}{b} e^{-\frac{z}{b}} dz \right) dy - g'(0) = \frac{1}{b^2} \int_0^\infty \int_0^z g'(y) e^{-\frac{z}{b}} dy dz - g'(0) \\
&= \frac{1}{b^2} \int_0^\infty (g(z) - g(0)) e^{-\frac{z}{b}} dz - g'(0) = \frac{1}{b^2} \int_0^\infty g(z) e^{-\frac{z}{b}} dz - \frac{g(0)}{b^2} \int_0^\infty e^{-\frac{z}{b}} dz - g'(0) \\
&= \frac{1}{b^2} \int_0^\infty g(z) e^{-\frac{z}{b}} dz - \frac{g(0)}{b} - g'(0).
\end{aligned}$$

(The fact that $E[g(W)]$ exists justifies the use of Fubini by reading the above backwards.)

Setting $h(y) = g(-y)$, it follows from the previous calculation that

$$\begin{aligned}
\int_{-\infty}^0 g''(x) e^{\frac{x}{b}} dx &= \int_0^\infty g''(-y) e^{-\frac{y}{b}} dy = \int_0^\infty h''(y) e^{-\frac{y}{b}} dy \\
&= \frac{1}{b^2} \int_0^\infty h(z) e^{-\frac{z}{b}} dz - \frac{h(0)}{b} - h'(0) \\
&= \frac{1}{b^2} \int_0^\infty g(-z) e^{-\frac{z}{b}} dz - \frac{g(0)}{b} + g'(0) \\
&= \frac{1}{b^2} \int_{-\infty}^0 g(z) e^{\frac{z}{b}} dz - \frac{g(0)}{b} + g'(0).
\end{aligned}$$

Therefore, since the preceding equations show that $\left| \int_{-\infty}^0 g''(x) e^{\frac{x}{b}} dx \right|, \left| \int_0^\infty g''(x) e^{-\frac{x}{b}} dx \right| < \infty$, we see that

$$\begin{aligned}
E[g''(W)] &= \frac{1}{2b} \int_{-\infty}^0 g''(x) e^{\frac{x}{b}} dx + \frac{1}{2b} \int_0^\infty g''(x) e^{-\frac{x}{b}} dx \\
&= \frac{1}{2b} \left[\left(\frac{1}{b^2} \int_{-\infty}^0 g(z) e^{\frac{z}{b}} dz - \frac{g(0)}{b} + g'(0) \right) + \left(\frac{1}{b^2} \int_0^\infty g(z) e^{-\frac{z}{b}} dz - \frac{g(0)}{b} - g'(0) \right) \right] \\
&= \frac{1}{2b} \left(\frac{1}{b^2} \int_{-\infty}^\infty g(z) e^{-\frac{|z|}{b}} dz - 2 \frac{g(0)}{b} \right) = \frac{1}{b^2} (E[g(W)] - g(0)).
\end{aligned}$$

□

Observe that if X is any random variable which has all of its moments and satisfies the equation in Lemma 1, then, taking $g(k) = x^k$, we have $E[X^k] = b^2 k(k-1)E[X^{k-2}]$ for all $k \in \mathbb{N}$. Since $E[X] = 0 + b^2 E[0] = 0$ and $E[X^2] = 0 + b^2 E[2] = 2b^2$, this recursion implies that

$$\mu_k := E[X^k] = \begin{cases} 0, & k \text{ is odd} \\ b^k k!, & k \text{ is even} \end{cases}$$

for all $k \in \mathbb{N}$, which is precisely the moment sequence for a $\text{Laplace}(0, b)$ random variable. Because $(\mu_{2k})^{\frac{1}{2k}} = b[(2k)!]^{\frac{1}{2k}} \leq b[(2k)^{2k}]^{\frac{1}{2k}} = 2bk$ for all $k \in \mathbb{N}$, we have $\sum_{k=1}^\infty \left(\frac{1}{\mu_{2k}} \right)^{\frac{1}{2k}} \geq \frac{1}{2b} \sum_{k=1}^\infty \frac{1}{k} = \infty$, so it follows from Carleman's criterion [5] that the $\text{Laplace}(0, b)$ distribution is the unique distribution having moments $\{\mu_k\}_{k=1}^\infty$. Therefore, the equation in Lemma 1 completely characterizes the Laplace distribution amongst all distributions for which every moment exists.

Note also that since the density of a $\text{Laplace}(0, b)$ random variable is given by $f_W(w) = \frac{1}{2b}e^{-\frac{|w|}{b}}$, the density method [3] suggests the following Stein equation for the Laplace distribution:

$$0 = g'(w) + \frac{f'_W(w)}{f_W(w)}g(w) = g'(w) - \frac{1}{b}\text{sgn}(w)g(w),$$

and, indeed, one can verify that if $W \sim \text{Laplace}(0, b)$, then

$$E[g'(W)] = \frac{1}{b}E[\text{sgn}(W)g(W)]$$

for all absolutely continuous g for which these expectations exist. Thus if g' is such a function as well, setting $G(w) = \text{sgn}(w)(g(w) - g(0))$, we have

$$E[g''(W)] = \frac{1}{b}E[\text{sgn}(W)g'(W)] = \frac{1}{b}E[G'(W)] = \frac{1}{b^2}E[\text{sgn}(W)G(W)] = \frac{1}{b^2}(E[g(W)] - g(0)),$$

so the general form of the equation in Lemma 1 can be ascertained by iterating the density method.

Now, in order to establish that the equation in Lemma 1 completely characterizes the $\text{Laplace}(0, b)$ distribution, we will show that if X is any random variable satisfying this equation, then

$$d_{BL}(\mathcal{L}(X), \text{Laplace}(0, b)) = 0$$

where d_{BL} denotes the bounded Lipschitz distance given by

$$d_{BL}(\mathcal{L}(X), \mathcal{L}(Y)) = \sup_{h \in \mathcal{H}_{BL}} |E[h(X)] - E[h(Y)]|,$$

$$\mathcal{H}_{BL} = \{h : \|h\|_\infty \leq 1 \text{ and } |h(x) - h(y)| \leq |x - y| \text{ for all } x, y \in \mathbb{R}\}.$$

As d_{BL} is a metric on the space of Borel probability measures on \mathbb{R} [19], this will establish the converse to Lemma 1. We begin by recording the following proposition relating d_{BL} to the more familiar Kolmogorov distance

$$d_K(\mathcal{L}(X), \mathcal{L}(Y)) = \sup_{x \in \mathbb{R}} |\mathbb{P}\{X \leq x\} - \mathbb{P}\{Y \leq x\}|.$$

Proposition 1. *If Z is an absolutely continuous random variable whose density, f_Z , is uniformly bounded by a constant $C < \infty$, then for any random variable X ,*

$$d_K(\mathcal{L}(X), \mathcal{L}(Z)) \leq \frac{C+2}{2} \sqrt{d_{BL}(\mathcal{L}(X), \mathcal{L}(Z))}.$$

Proof. We first note that the inequality holds trivially if $d_{BL}(\mathcal{L}(X), \mathcal{L}(Z)) = 0$ since d_{BL} and d_K are metrics. Also, since $d_K(P, Q) \leq 1$ for all probability measures P and Q , $\frac{C+2}{2} \geq 1$, and $d_{BL}(\mathcal{L}(X), \mathcal{L}(Z)) \geq 1$ implies $\sqrt{d_{BL}(\mathcal{L}(X), \mathcal{L}(Z))} \geq 1$, we have

$$d_K(\mathcal{L}(X), \mathcal{L}(Z)) \leq 1 \leq \frac{C+2}{2} \sqrt{d_{BL}(\mathcal{L}(X), \mathcal{L}(Z))}$$

whenever $d_{BL}(\mathcal{L}(X), \mathcal{L}(Z)) \geq 1$. As such, we may suppose that $d_{BL}(\mathcal{L}(X), \mathcal{L}(Z)) \in (0, 1)$.

Now, for $x \in \mathbb{R}$, $\varepsilon > 0$, write

$$h_x(z) = 1_{(-\infty, x]}(z) = \begin{cases} 1, & z \leq x \\ 0, & z > x \end{cases} \quad \text{and} \quad h_{x, \varepsilon}(z) = \begin{cases} 1, & z \leq x \\ 1 - \frac{z-x}{\varepsilon}, & z \in (x, x+\varepsilon) \\ 0, & z > x+\varepsilon \end{cases}.$$

Then for all $x \in \mathbb{R}$,

$$\begin{aligned}
E[h_x(X) - h_x(Z)] &= E[h_x(X)] - E[h_{x,\varepsilon}(Z)] + E[h_{x,\varepsilon}(Z)] - E[h_x(Z)] \\
&\leq (E[h_{x,\varepsilon}(X)] - E[h_{x,\varepsilon}(Z)]) + \int_x^{x+\varepsilon} \left(1 - \frac{z-x}{\varepsilon}\right) f_Z(z) dz \\
&\leq |E[h_{x,\varepsilon}(X)] - E[h_{x,\varepsilon}(Z)]| + \frac{C\varepsilon}{2}.
\end{aligned}$$

Since $d_{BL}(\mathcal{L}(X), \mathcal{L}(Z)) \in (0, 1)$, if we take $\varepsilon = \sqrt{d_{BL}(\mathcal{L}(X), \mathcal{L}(Z))} \in (0, 1)$, then $\varepsilon h_{x,\varepsilon} \in \mathcal{H}_{BL}$ and thus

$$\begin{aligned}
E[h_x(X) - h_x(Z)] &\leq \frac{1}{\varepsilon} |E[\varepsilon h_{x,\varepsilon}(X)] - E[\varepsilon h_{x,\varepsilon}(Z)]| + \frac{C\varepsilon}{2} \\
&\leq \frac{1}{\varepsilon} d_{BL}(\mathcal{L}(X), \mathcal{L}(Z)) + \frac{C\varepsilon}{2} \\
&= \frac{C+2}{2} \sqrt{d_{BL}(\mathcal{L}(X), \mathcal{L}(Z))}.
\end{aligned}$$

A similar argument using

$$\begin{aligned}
E[h_x(Z) - h_x(X)] &= E[h_x(Z)] - E[h_{x-\varepsilon,\varepsilon}(Z)] + E[h_{x-\varepsilon,\varepsilon}(Z)] - E[h_x(X)] \\
&\leq \frac{C\varepsilon}{2} + (E[h_{x-\varepsilon,\varepsilon}(Z)] - E[h_{x-\varepsilon,\varepsilon}(X)])
\end{aligned}$$

shows that

$$|E[h_x(X)] - E[h_x(Z)]| \leq \frac{C+2}{2} \sqrt{d_{BL}(\mathcal{L}(X), \mathcal{L}(Z))}$$

for all $x \in \mathbb{R}$, and the proposition follows by taking suprema. \square

Remark. When $C \geq 1$, we can take $\varepsilon = \sqrt{\frac{1}{C} d_{BL}(\mathcal{L}(X), \mathcal{L}(Z))}$ in the above argument to obtain an improved bound of

$$d_K(\mathcal{L}(X), \mathcal{L}(Z)) \leq \frac{3}{2} \sqrt{C d_{BL}(\mathcal{L}(X), \mathcal{L}(Z))}.$$

To the best of the authors' knowledge, the above proposition is original, though the proof follows the same basic line of reasoning as the well-known bound on the Kolmogorov distance by the Wasserstein distance (see Proposition 1.2 in [15]). It seems that the primary reason for using the Wasserstein metric, d_W , is that it enables one to work with smoother functions while still implying convergence in the more natural Kolmogorov distance. Proposition 1 shows that d_{BL} also upper-bounds d_K while enjoying all of the resulting smoothness of d_W and with additional boundedness properties to boot. Also, d_{BL} is a fairly natural measure of distance since it metrizes weak convergence [19]. However, convergence in d_W is always stronger than convergence in d_{BL} and it is possible to have $d_{BL}(\mathcal{L}(X), \mathcal{L}(Z)) = 0$ while $d_K(\mathcal{L}(X), \mathcal{L}(Z)) > 0$ when the distribution of Z does not satisfy the assumptions of Proposition 1. Furthermore, the bounded Lipschitz metric does not scale as nicely as the Kolmogorov or Wasserstein distances when the associated random variables are multiplied by a positive constant. For the remainder of this paper, we will state our results in terms of d_{BL} with the corresponding Kolmogorov bound being implicit therein, though one should note that, as with the Wasserstein bound on d_K , Kolmogorov bounds obtained in this fashion are not necessarily optimal and often give the root of the true rate.

Now, in light of Lemma 1, we are led to consider the solution to the initial value problem

$$g(x) - b^2 g''(x) = h(x) - Wh, \quad g(0) = 0$$

where $h \in \mathcal{H}_{BL}$ and $Wh = E[h(W)]$, $W \sim \text{Laplace}(0, b)$. For the sake of compactness, we will write $\tilde{h}(x) := h(x) - Wh$.

The general solution to the homogeneous equation $g''(x) - b^{-2}g(x) = 0$ is given by $g_0(x) = C_1 e^{\frac{x}{b}} + C_2 e^{-\frac{x}{b}}$, so, since $\begin{vmatrix} e^{\frac{x}{b}} & e^{-\frac{x}{b}} \\ \frac{1}{b}e^{\frac{x}{b}} & -\frac{1}{b}e^{-\frac{x}{b}} \end{vmatrix} = -\frac{2}{b} \neq 0$, it follows from the variation of parameters method that a solution to the inhomogeneous equation $g''(x) - b^{-2}g(x) = -b^{-2}\tilde{h}(x)$ is given by

$$g_h(x) = v_h^1(x)e^{\frac{x}{b}} + v_h^2(x)e^{-\frac{x}{b}}$$

where

$$v_h^1(x) = \frac{1}{2b} \int_x^\infty e^{-\frac{y}{b}} \tilde{h}(y) dy \text{ and } v_h^2(x) = \frac{1}{2b} \int_{-\infty}^x e^{\frac{y}{b}} \tilde{h}(y) dy.$$

To see that the initial condition is satisfied, we observe that

$$\begin{aligned} g_h(0) &= v_h^1(0) + v_h^2(0) = \frac{1}{2b} \int_{-\infty}^\infty e^{-\frac{|y|}{b}} \tilde{h}(y) dy = \int_{-\infty}^\infty f_W(y) (h(y) - Wh) dy \\ &= \int_{-\infty}^\infty h(y) f_W(y) dy - Wh \int_{-\infty}^\infty f_W(y) dy = Wh - Wh = 0. \end{aligned}$$

Taking derivatives yields

$$\begin{aligned} g_h'(x) &= v_h^{1'}(x)e^{\frac{x}{b}} + \frac{1}{b}v_h^1(x)e^{\frac{x}{b}} + v_h^{2'}(x)e^{-\frac{x}{b}} - \frac{1}{b}v_h^2(x)e^{-\frac{x}{b}} \\ &= \frac{1}{2b}\tilde{h}(x) + \frac{1}{b}v_h^1(x)e^{\frac{x}{b}} - \frac{1}{2b}\tilde{h}(x) - \frac{1}{b}v_h^2(x)e^{-\frac{x}{b}} = \frac{1}{b}(v_h^1(x)e^{\frac{x}{b}} - v_h^2(x)e^{-\frac{x}{b}}), \\ g_h''(x) &= \frac{1}{b^2}(-\tilde{h}(x) + v_h^1(x)e^{\frac{x}{b}} + v_h^2(x)e^{-\frac{x}{b}}), \end{aligned}$$

and

$$g_h'''(x) = -\frac{1}{b^2}h'(x) + \frac{1}{b^3}(v_h^1(x)e^{\frac{x}{b}} - v_h^2(x)e^{-\frac{x}{b}}).$$

(The expression for g_h'' may also be obtained by noting that g_h satisfies $g_h(x) - b^2 g_h''(x) = \tilde{h}(x)$ and the expression for g_h''' then follows by differentiating $g_h''(x) = \frac{1}{b^2}(g_h(x) - h(x) + Wh)$.)

Moreover, as $\|h\|_\infty \leq 1$,

$$|Wh| = \left| \frac{1}{2b} \int_{-\infty}^\infty h(x) e^{-\frac{|x|}{b}} dx \right| \leq \frac{1}{2b} \int_{-\infty}^\infty e^{-\frac{|x|}{b}} dx = 1,$$

and thus $|\tilde{h}(x)| \leq |h(x)| + |Wh| \leq 2$. Consequently,

$$|v_h^1(x)e^{\frac{x}{b}}| \leq \frac{1}{2b}e^{\frac{x}{b}} \int_x^\infty 2e^{-\frac{y}{b}} dy = 1$$

and

$$|v_h^2(x)e^{-\frac{x}{b}}| \leq \frac{1}{2b}e^{-\frac{x}{b}} \int_{-\infty}^x 2e^{\frac{y}{b}} dy = 1,$$

for all $x \in \mathbb{R}$, so, since $\|h'\|_\infty \leq 1$ as well, we have

Lemma 2. For $h \in \mathcal{H}_{BL}$, $W \sim \text{Laplace}(0, b)$, $\tilde{h}(x) := h(x) - Wh$, a bounded solution to the initial value problem

$$g(x) - b^2 g''(x) = \tilde{h}(x), \quad g(0) = 0$$

is given by

$$g_h(x) = \frac{1}{2b} \left(e^{\frac{x}{b}} \int_x^\infty e^{-\frac{y}{b}} \tilde{h}(y) dy + e^{-\frac{x}{b}} \int_{-\infty}^x e^{\frac{y}{b}} \tilde{h}(y) dy \right).$$

This solution satisfies $\|g_h\|_\infty \leq 2$, $\|g'_h\|_\infty \leq \frac{2}{b}$, $\|g''_h\|_\infty \leq \frac{4}{b^2}$ and $\|g'''_h\|_\infty \leq \frac{b+2}{b^3}$. Consequently, $E|g_h(X)|$, $E|g'_h(X)|$, and $E|g''_h(X)|$ exist for all X , g_h and g'_h are absolutely continuous, and g''_h is Lipschitz with Lipschitz constant $\frac{b+2}{b^3}$.

With the preceding results in hand, we are in a position to establish that the equation from Lemma 1 does indeed characterize the Laplace distribution.

Lemma 3. *If X is a random variable such that*

$$E[g(X)] - g(0) = b^2 E[g''(X)]$$

for every absolutely continuous function g such that g' is absolutely continuous and $E[g(X)]$ exists, then $X \sim \text{Laplace}(0, b)$.

Proof. Let $W \sim \text{Laplace}(0, b)$ and, for $h \in \mathcal{H}_{BL}$, let g_h be as in Lemma 2. Because $g_h(0) = 0$ and g_h, g'_h, g''_h are bounded, it follows from the above assumptions that

$$E[h(X)] - E[h(W)] = E[g_h(X) - b^2 g''_h(X)] = 0.$$

Taking suprema over $h \in \mathcal{H}_{BL}$ shows that $d_{BL}(\mathcal{L}(X), \mathcal{L}(W)) = 0$. □

Therefore, $X \sim \text{Laplace}(0, b)$ if and only if $E[g(X)] - g(0) = b^2 E[g''(X)]$ for every absolutely continuous function g such that g' is absolutely continuous and $E[g(X)]$ exists.

3. BOUNDING THE ERROR

Our next task is to use this characterization to obtain bounds on the error terms resulting from approximation by the Laplace distribution. To this end, we introduce the following definition: For any random variable X with finite second moment, we say that the random variable X^L has the *symmetric equilibrium distribution* with respect to X if

$$E[f(X)] - f(0) = \frac{1}{2} E[X^2] E[f''(X^L)]$$

for all functions f such that f and f' are absolutely continuous and $E|f(X)|, E|f'(X)| < \infty$. We call the map $X \mapsto X^L$ the symmetric equilibrium transformation. (The nomenclature is based on similarities with the equilibrium distribution from renewal theory which was used in a similar manner in [12] for a related problem involving the exponential distribution.)

The utility of this definition is that if X is a random variable which is in the domain of the symmetric equilibrium transformation, then we can bound the distance between the law of X and the Laplace distribution with the following theorem.

Theorem 1. *If X is a random variable such that $E[X^2] = 2b^2$, and X^L exists, then*

$$d_{BW}(\mathcal{L}(X), \text{Laplace}(0, b)) \leq \frac{b+2}{b} E|X - X^L|.$$

Proof. If X satisfies the assumptions of the theorem and X^L has the symmetric equilibrium distribution for X , then for all $h \in \mathcal{H}_{BL}$, taking g_h as above, we see that

$$\begin{aligned} |Wh - E[h(X)]| &= |E[g_h(X) - b^2 g_h''(X)]| = |E[b^2 g_h''(X^L) - b^2 g_h''(X)]| \\ &\leq b^2 E |g_h''(X^L) - g_h''(X)| \leq b^2 \|g_h''\| E |X^L - X| \\ &= \frac{b+2}{b} E |X - X^L|. \end{aligned}$$

□

In addition, we have the complementary result

Proposition 2. *If Y^L has the symmetric equilibrium distribution for Y where $E[Y^2] = 2b^2$, then*

$$E |Y - Y^L| \leq E |Y| + \frac{1}{6b^2} E[|Y|^3].$$

Proof. The inequality holds trivially if $E[|Y|^3] = \infty$. Otherwise, we have

$$E |Y - Y^L| \leq (E |Y| + E |Y^L|) = E |Y| + \frac{1}{6b^2} E[|Y|^3],$$

where the equality follows upon taking $f(y) = |y|^3$ in the definition of the transformation $Y \mapsto Y^L$. □

At this point, we need to show that the domain of the symmetric equilibrium transformation is large enough for these results to be of use. In order to do so, we appeal to the following theorem due to Larry Goldstein and Gesine Reinert concerning generalizations of the size bias and zero bias transformations [7].

Theorem 2 (Goldstein and Reinert). *Let X be a random variable, $m \in \{0, 1, 2, \dots\}$, and P a measurable function with exactly m sign changes, positive on its rightmost interval, such that*

$$\frac{1}{m!} E[X^k P(X)] = \alpha \delta_{k,m}, \quad k = 0, 1, \dots, m$$

for some $\alpha > 0$. Then there exists a unique distribution for a random variable $X^{(P)}$ such that

$$E[P(X)F(X)] = \alpha E[F^{(m)}(X^{(P)})]$$

for all $F \in \mathcal{F}^m(X, P) = \{F : F^{(m)} \text{ exists and is measurable, } E |P(X)F(X)| < \infty\}$. We say that $X^{(P)}$ has the $X - P$ biased distribution.

Remark. The proof of Theorem 2 shows that in the definition of $\mathcal{F}^m(X, P)$, it suffices to assume that $F^{(m-1)}$ is absolutely continuous and $F^{(m)}$ exists on a set S with $\mathbb{P}\{X \in S\} = 1$, and we will assume this in what follows. Thus, for example, the absolute value function lies in $\mathcal{F}^1(X, P)$ as long as $E |XP(X)| < \infty$.

The basic idea of our existence argument is to iterate the mapping $X \mapsto X^{(P)}$ for a suitable function P in much the same way as we showed that we could derive the second order Stein equation for the Laplace distribution by iterating the density method. The idea of composing distributional transformations appears in [13] as well, and their results also involve a second order characterizing equation. The precise statement of our result is

Theorem 3. *Let X be any random variable satisfying $E[X] = 0$, $\mathbb{P}\{X < 0\} = \mathbb{P}\{X > 0\} = \frac{1}{2}$, and $E[X^2] < \infty$. Then there exists a random variable X^L which satisfies $E[f(X)] - f(0) = \frac{1}{2} E[X^2] E[f''(X^L)]$ for all functions f such that f, f' are absolutely continuous and $E |f(X)|, E |f'(X)|$ exist.*

Proof. Taking $P(x) = \text{sgn}(X) = \begin{cases} -1, & x < 0 \\ 1, & x \geq 0 \end{cases}$, it follows from the assumptions on X that

$$\frac{1}{0!}E[X^0 P(X)] = E[\text{sgn}(X)] = 0 \text{ and } \frac{1}{1!}E[X^1 \text{sgn}(X)] = E|X| := \alpha > 0,$$

thus, by Theorem 2, there exists a distribution for a random variable $X^{(P)}$ with $E[P(X)F(X)] = \alpha E[F'(X^{(P)})]$ for all $F \in \mathcal{F}^1(X, P)$.

Since $\frac{d}{dx}|x| = \text{sgn}(x)$ for $x \neq 0$, and $\mathbb{P}\{X = 0\} = 0$ by assumption, it follows from the definition of $X^{(P)}$ that

$$E[\text{sgn}(X^{(P)})] = \frac{1}{\alpha}E[\text{sgn}(X)|X|] = \frac{1}{\alpha}E[X] = 0,$$

and similarly, since $\frac{d}{dx}\left(\frac{x^2}{2}\text{sgn}(x)\right) = |x|$,

$$E[X^{(P)}\text{sgn}(X^{(P)})] = E|X^{(P)}| = \frac{1}{\alpha}E[\text{sgn}(X)\frac{1}{2}X^2\text{sgn}(X)] = \frac{1}{2\alpha}E[X^2] := \beta.$$

As such, Theorem 2 asserts the existence of a random variable, which we will call X^L , that has the $X^{(P)} - P$ biased distribution.

Therefore, if f is any function satisfying the assumptions in the statement of the theorem, setting $g(x) = \text{sgn}(x)(f(x) - f(0)) \in \mathcal{F}^1(X, P)$, we have

$$\begin{aligned} E[f(X)] - f(0) &= E[\text{sgn}(X)g(X)] = \alpha E[g'(X^{(P)})] = \alpha E[\text{sgn}(X^{(P)})f'(X^{(P)})] \\ &= \alpha\beta E[f''(X^L)] = \frac{1}{2}E[X^2]E[f''(X^L)]. \end{aligned}$$

□

Theorem 3 can also be derived directly from the Riesz representation theorem using the same general argument, but Theorem 2 was the original inspiration and it spares us some of the technical details involved in a direct proof.

It is worth mentioning that the choice of $P(x) = \text{sgn}(x)$ for constructing a transformation suited to the Laplace(0, b) distribution has connections with the density method for constructing an appropriate Stein equation in that $\frac{f'_W(x)}{f_W(x)} = -\frac{1}{b}\text{sgn}(x)$. This connection is also evident in the zero bias transformation which is used in normal approximation. In general, if one is trying to construct a first order transformation suited to a random variable with density $p(x)$, then a good initial guess for choosing $P(x)$ in the definition of the $X - P$ biased transformation is to take $P(x) = -\frac{p'(x)}{p(x)}$ (or some constant multiple thereof). This makes sense because if $E[F'(X) + \frac{p'(X)}{p(X)}F(X)] = 0$, then X is a fixed point of the transformation $X \mapsto X^{(P)}$ with $E[-\alpha\frac{p'(X)}{p(X)}F(X)] = \alpha E[F'(X^{(P)})]$. The present work shows that it may sometimes be useful to iterate these heuristics to construct higher order Stein equations and distributional transformations.

We conclude this section by mentioning some properties of the symmetric equilibrium distribution. Here X is taken to be a random variable in the domain of the symmetric equilibrium transformation which has mean 0 and variance $2b^2$. To begin with, if X has moments of all orders, then by considering the function $f(x) = x^{k+2}$, we see that the moments of X^L are given by

$$E[(X^L)^k] = \frac{\mu_{k+2}}{b^2(k+2)(k+1)}, \quad \mu_k = E[X^k].$$

A direct computation using $E[e^{itx}] = \sum_{k=0}^{\infty} \frac{(it)^k}{k!} \mu_k$ then shows that X^L has characteristic function $\varphi_{X^L}(t) = \frac{1 - \varphi_X(t)}{t^2 b^2}$ where φ_X is the ch.f. of X . This can also be ascertained without all the assumptions on the moments of X by taking $g_t(x) = -t^{-2}e^{itx}$ in the definition of X^L .

In addition, since the proof of Theorem 2 is constructive, one can determine the form of X^L without resorting to characteristic function arguments. Specifically, using the notation of Theorem 3, one can show that $X^{(P)}$ is of the form $U_1 Y$ where U_1 and Y are independent with $U_1 \sim \text{Uniform}(0, 1)$ and Y having density $f_Y(y) = \frac{|y|}{E|X|} f_X(y)$, and X_L is of the form $U_2 Z$ where U_2 and Z are independent with $U_2 \sim \text{Uniform}(0, 1)$ and Z having density $f_Z(z) = \frac{|z|}{E|X^{(P)}|} f_{X^{(P)}}(z) = \frac{2E|X|}{E[X^2]} |z| f_{X^{(P)}}(z)$ [7]. Consequently, we may write the density of X^L as

$$f_{X^L}(s) = \frac{s^2}{b^2} \int_0^1 \int_0^1 \frac{1}{u^2 v^3} f_x\left(\frac{s}{uv}\right) dudv.$$

Finally, since the characterizing equation involves second derivatives and a variance term, one expects some kind of relation between X^L and the zero bias distribution for X (defined by $E[Xf(X)] = E[X^2]E[f'(X^z)]$ for all absolutely continuous f for which these expectations exist), in much the same way as the equilibrium distribution is related to the size bias distribution [12]. Though we cannot express X^L directly in terms of X^z , a similar argument shows that if $U \sim \text{Uniform}(0, 1)$ is independent of all else, then

$$\begin{aligned} \frac{1}{2}E[f''(X^L)] &= \frac{1}{E[X^2]} (E[f(X)] - f(0)) = \frac{1}{E[X^2]} E\left[\int_0^1 X f'(uX) du\right] \\ &= \int_0^1 \frac{1}{E[X^2]} E[X f'(uX)] du = \int_0^1 E[uf''(uX^z)] du = E[U f''(UX^z)] \end{aligned}$$

for all functions f such that f, f' are absolutely continuous and the above expectations exist. We have not been able to make much use of the preceding results in our investigations of geometric sums, but they may come in handy for other applications involving the Laplace distribution.

4. RANDOM SUMS

The *p-geometric summation* of a sequence of random variables X_1, X_2, \dots is defined as $S_p = X_1 + X_2 + \dots + X_{N_p}$ where N_p is geometric with success probability p - that is, $\mathbb{P}\{N_p = n\} = p(1-p)^{n-1}$, $n \in \mathbb{N}$ - and is independent of all else. A result due to Rényi [14] states that if X_1, X_2, \dots are i.i.d., positive, nondegenerate random variables with $E[X_i] = 1$, then $\mathcal{L}(pS_p) \rightarrow \text{Exponential}(1)$ as $p \rightarrow 0$. In fact, just as the normal law is the only nondegenerate distribution with finite variance which is stable under ordinary summation (in the sense that if X, X_1, X_2, \dots are i.i.d. nondegenerate random variables with finite variance, then for every $n \in \mathbb{N}$, there exist $a_n > 0$, $b_n \in \mathbb{R}$ such that $X \stackrel{d}{=} a_n(X_1 + \dots + X_n) + b_n$), it can be shown that if X, X_1, X_2, \dots are i.i.d., positive, and nondegenerate with finite variance, then there exists $a_p \in \mathbb{R}$ such that $a_p(X_1 + \dots + X_{N_p}) \stackrel{d}{=} X$ for all $p \in (0, 1)$ if and only if X has an exponential distribution. Similarly, if we assume that Y, Y_1, Y_2, \dots are i.i.d., symmetric, and nondegenerate with finite variance, then there exists $a_p \in \mathbb{R}$ $a_p(Y_1 + \dots + Y_{N_p}) \stackrel{d}{=} Y$ for all $p \in (0, 1)$ if and only if Y has a Laplace distribution. Moreover, it must be the case that $a_p = p^{\frac{1}{2}}$. In addition, we have an analog of Rényi's Theorem [9]:

Theorem 4. Suppose that X_1, X_2, \dots are i.i.d. symmetric and nondegenerate random variables with finite variance σ^2 , and let $N_p \sim \text{Geometric}(p)$ be independent of the X_i 's. If

$$a_p \sum_{i=1}^{N_p} X_i \rightarrow_d X \text{ as } p \rightarrow 0,$$

then there exists $\gamma > 0$ such that $a_p = p^{\frac{1}{2}}\gamma + o(p^{\frac{1}{2}})$ and X has the Laplace distribution with mean 0 and variance $\sigma^2\gamma^2$.

A recent theorem due to Alexis Toda gives the following Lindeberg-type conditions for the existence of the distributional limit in Theorem 4.

Theorem 5 (Toda). *Let X_1, X_2, \dots be a sequence of independent (but not necessarily identically distributed) random variables such that $E[X_i] = 0$ and $\text{Var}(X_i) = \sigma_i^2$, and let $N_p \sim \text{Geometric}(p)$ independent of the X_i 's. Suppose that*

$$\lim_{n \rightarrow \infty} n^{-\alpha} \sigma_n^2 = 0 \text{ for some } 0 < \alpha < 1,$$

$$\sigma^2 := \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \sigma_i^2 > 0 \text{ exists,}$$

and for all $\epsilon > 0$,

$$\lim_{p \rightarrow 0} \sum_{i=1}^{\infty} (1-p)^{i-1} p E[X_i^2; |X_i| \geq \epsilon p^{-\frac{1}{2}}] = 0.$$

Then as $p \rightarrow 0$, the sum $p^{\frac{1}{2}} \sum_{i=1}^{N_p} X_i$ converges weakly to the Laplace distribution with mean 0 and variance σ^2 .

Remark. The original statement of Toda's Theorem is slightly more general, allowing for convergence to a possibly asymmetric Laplace distribution, but we will ignore that version for the moment as we state the result only for the sake of comparison.

In 2011, Peköz and Röllin were able to generalize Rényi's theorem by using a distributional transformation inspired by Stein's method considerations [12]. Specifically, for a nonnegative random variable X with $E[X] < \infty$, they say that X^e has the equilibrium distribution with respect to X if $E[f(X)] - f(0) = E[X]E[f'(X^e)]$ for all Lipschitz f and use this to bound the Wasserstein and Kolmogorov distances of a distribution to the Exponential(1) distribution. (The equilibrium distribution arises in renewal theory, but its utility in analyzing convergence to the exponential distribution comes from the fact that a Stein operator for the exponential distribution with mean one is given by $(\mathcal{A}_E f)(x) = f'(x) - f(x) + f(0)$, so $X \sim \text{Exponential}(1)$ is a fixed point of the equilibrium transformation [15].) This transformation and the similarity between our Stein characterization of the Laplace and the above characterization of the exponential inspired our construction of the symmetric equilibrium transformation, and the fact that both distributions are stable under geometric summation led us to parallel their argument for bounding the distance between p -geometric sums of positive random variables and the exponential distribution in order to obtain corresponding results for the Laplace case. Our results are summarized in the following theorem and its corollaries.

Theorem 6. *Let N be any \mathbb{N} -valued random variable with $\mu = E[N] < \infty$ and let X_1, X_2, \dots be a sequence of independent random variables, independent of N , with $\mathbb{P}\{X_i < 0\} = \mathbb{P}\{X_i > 0\} = \frac{1}{2}$, $E[X_i] = 0$, and $E[X_i^2] = \sigma_i^2 < \infty$. Set $\sigma^2 = E\left[\left(\sum_{i=1}^N X_i\right)^2\right] = E\left[\sum_{i=1}^N \sigma_i^2\right]$ and let M be any \mathbb{N} -valued random variable, defined on the same space as N , satisfying*

$$\mathbb{P}\{M = m\} = \frac{\sigma_m^2}{\sigma^2} \mathbb{P}\{N \geq m\}.$$

Then

$$\begin{aligned}
& d_{BL} \left(\mathcal{L}(\mu^{-\frac{1}{2}} \sum_{i=1}^N X_i), \text{Laplace} \left(0, \frac{\sigma}{\sqrt{2\mu}} \right) \right) \\
& \leq \left(\mu^{-\frac{1}{2}} + \frac{\sqrt{8}}{\sigma} \right) E \left| (X_M - X_M^L) + \text{sgn}(N - M) \sum_{i=(N \wedge M)+1}^{N \vee M} X_i \right| \\
& \leq \left(\mu^{-\frac{1}{2}} + \frac{\sqrt{8}}{\sigma} \right) \left(E |X_M| + \frac{1}{3} E \left[\frac{1}{\sigma_M^2} |X_M|^3 \right] + \sup_{i \geq 1} \sigma_i E[|N - M|^{\frac{1}{2}}] \right).
\end{aligned}$$

Proof. We first note that

$$\sigma^2 = \sum_{m=1}^{\infty} \mathbb{P}\{N = m\} \sum_{i=1}^m \sigma_i^2 = \sum_{m=1}^{\infty} \mathbb{P}\{N \geq m\} \sigma_m^2,$$

so M is well-defined.

Now, taking $V = \mu^{-\frac{1}{2}} \sum_{i=1}^N X_i$, we claim that $V^L = \mu^{-\frac{1}{2}} \left(\sum_{i=1}^{M-1} X_i + X_M^L \right)$ has the symmetric equilibrium distribution with respect to V . To see this, let f be any function satisfying the assumptions of Theorem 3. Then, using the notation

$$g(m) = f(\mu^{-\frac{1}{2}} \sum_{i=1}^m X_i), \quad h_s(x) = f\left(\mu^{-\frac{1}{2}} s + \mu^{-\frac{1}{2}} x\right),$$

letting ν_m denote the distribution of

$$S_{m-1} := \sum_{i=1}^{m-1} X_i,$$

and observing that, by independence,

$$E[h''(X_m^L) | S_{m-1} = s] = E[h''(X_m^L)] = \frac{2}{\sigma_m^2} (E[h(X_m) - h(0)]) = \frac{2}{\sigma_m^2} E[h(X_m) - h(0) | S_{m-1} = s],$$

for all suitable functions h , we see that

$$\begin{aligned}
E \left[f''(\mu^{-\frac{1}{2}} \sum_{i=1}^{m-1} X_i + \mu^{-\frac{1}{2}} X_m^L) \right] &= \int E \left[f''(\mu^{-\frac{1}{2}} s + \mu^{-\frac{1}{2}} X_m^L) | S_{m-1} = s \right] d\nu_m(s) \\
&= \int E [\mu h_s''(X_m^L) | S_{m-1} = s] d\nu_m(s) \\
&= \frac{2\mu}{\sigma_m^2} \int E [h_s(X_m) - h_s(0) | S_{m-1} = s] d\nu_m(s) \\
&= \frac{2\mu}{\sigma_m^2} \int E \left[f(\mu^{-\frac{1}{2}} s + \mu^{-\frac{1}{2}} X_m) - f(\mu^{-\frac{1}{2}} s) | S_{m-1} = s \right] d\nu_m(s) \\
&= \frac{2\mu}{\sigma_m^2} E \left[f(\mu^{-\frac{1}{2}} \sum_{i=1}^m X_i) - f(\mu^{-\frac{1}{2}} \sum_{i=1}^{m-1} X_i) \right] \\
&= \frac{2\mu}{\sigma_m^2} E [g(m) - g(m-1)]
\end{aligned}$$

for all $m \in \mathbb{N}$, hence

$$\begin{aligned}
E[f''(V^L)] &= \sum_{m=1}^{\infty} \mathbb{P}\{M = m\} E[f''(\mu^{-\frac{1}{2}} \sum_{i=1}^{m-1} X_i + \mu^{-\frac{1}{2}} X_m^L)] \\
&= \frac{2\mu}{\sigma^2} \sum_{m=1}^{\infty} \frac{\sigma^2}{\sigma_m^2} \mathbb{P}\{M = m\} E[g(m) - g(m-1)] \\
&= \frac{2}{E[V^2]} E \left[\sum_{m=1}^{\infty} \mathbb{P}\{N \geq m\} (g(m) - g(m-1)) \right] \\
&= \frac{2}{E[V^2]} (E[g(N)] - g(0)) = \frac{2}{E[V^2]} (E[f(V)] - f(0)).
\end{aligned}$$

Consequently, setting $2B^2 = E[V^2] = \frac{\sigma^2}{\mu}$, it follows from Theorem 1 that

$$\begin{aligned}
d_{BL}(\mathcal{L}(V), \text{Laplace}(0, B)) &\leq \left(1 + \frac{2}{B}\right) E|V - V^L| \\
&= \left(\mu^{-\frac{1}{2}} + \frac{\sqrt{8}}{\sigma}\right) E \left| (X_M - X_M^L) + \text{sgn}(N - M) \sum_{i=(N \wedge M)+1}^{N \vee M} X_i \right| \\
&\leq \left(\mu^{-\frac{1}{2}} + \frac{\sqrt{8}}{\sigma}\right) \left(E|X_M - X_M^L| + E \left| \sum_{i=(N \wedge M)+1}^{N \vee M} X_i \right| \right).
\end{aligned}$$

By Proposition 2, we have

$$\begin{aligned}
E|X_M - X_M^L| &= \sum_{m=1}^{\infty} \mathbb{P}\{M = m\} E|X_m - X_m^L| \\
&\leq \sum_{m=1}^{\infty} \mathbb{P}\{M = m\} \left(E|X_m| + \frac{1}{3\sigma_m^2} E[|X_m|^3] \right) \\
&= E|X_M| + \frac{1}{3} E \left[\frac{1}{\sigma_M^2} |X_M|^3 \right],
\end{aligned}$$

and, since the X_i 's are independent, it follows from the Cauchy-Schwarz inequality that

$$\begin{aligned}
E \left| \sum_{i=(N \wedge M)+1}^{N \vee M} X_i \right| &= \sum_{j=1}^{\infty} \sum_{k=1}^{\infty} \mathbb{P}\{N \wedge M = j, |N - M| = k\} E \left| \sum_{i=j+1}^{j+k} X_i \right| \\
&\leq \sum_{j=1}^{\infty} \sum_{k=1}^{\infty} \mathbb{P}\{N \wedge M = j, |N - M| = k\} E \left[\sum_{i=j+1}^{j+k} X_i^2 \right]^{\frac{1}{2}} \\
&\leq \sum_{j=1}^{\infty} \sum_{k=1}^{\infty} \mathbb{P}\{N \wedge M = j, |N - M| = k\} k^{\frac{1}{2}} \sup_{i \geq 1} \sigma_i \\
&= \sup_{i \geq 1} \sigma_i \sum_{k=1}^{\infty} k^{\frac{1}{2}} \mathbb{P}\{|N - M| = k\} = \sup_{i \geq 1} \sigma_i E[|N - M|^{\frac{1}{2}}].
\end{aligned}$$

□

Specializing to the i.i.d. case, so that $\sigma^2 = \mu\sigma_1^2$, we have the immediate corollary

Corollary 1. *Let N be any \mathbb{N} -valued random variable with $\mu = E[N] < \infty$ and let X_1, X_2, \dots be a sequence of i.i.d. random variables, independent of N , with $\mathbb{P}\{X_1 < 0\} = \mathbb{P}\{X_1 > 0\} = \frac{1}{2}$, $E[X_1] = 0$, $E[X_1^2] = 2b^2$, and $E[|X_1|^3] = \rho < \infty$. Let M be any \mathbb{N} -valued random variable, defined on the same space as N , satisfying*

$$\mathbb{P}\{M = m\} = \mu^{-1} \mathbb{P}\{N \geq m\}.$$

Then

$$d_{BL} \left(\mathcal{L}(\mu^{-\frac{1}{2}} \sum_{i=1}^N X_i), \text{Laplace}(0, b) \right) \leq \frac{b+2}{b\sqrt{\mu}} \left(E|X_1| + \frac{\rho}{6b^2} + b\sqrt{2}E[|N-M|^{\frac{1}{2}}] \right).$$

When $N \sim \text{Geometric}(p)$ and the X_i 's satisfy the assumptions of Theorem 6 along with the additional constraint $E[X_i^2] = 2b^2$ for all $i \geq 1$, we have $\mu = \frac{1}{p}$ and

$$\frac{\sigma_m^2}{\sigma^2} \mathbb{P}\{N \geq m\} = p \mathbb{P}\{N \geq m\} = p^2 \sum_{i=m}^{\infty} (1-p)^{m-1} = p(1-p)^{m-1} = \mathbb{P}\{N = m\},$$

so we can take $M = N$ in Theorem 6 to obtain

$$\begin{aligned} d_{BL} \left(\mathcal{L}(p^{\frac{1}{2}} \sum_{i=1}^N X_i), \text{Laplace}(0, b) \right) &\leq p^{\frac{1}{2}} \frac{b+2}{b} E|X_N - X_N^L| \\ &\leq p^{\frac{1}{2}} \frac{b+2}{b} \left(E|X_N| + \frac{1}{3} E \left[\frac{1}{\sigma_N^2} |X_N|^3 \right] \right). \end{aligned}$$

Consequently, using Cauchy-Schwarz to bound $E|X_N|$, we have the following Berry-Esseen type theorem.

Theorem 7. *Let X_1, X_2, \dots be a sequence of independent random variables with $\mathbb{P}\{X_i < 0\} = \mathbb{P}\{X_i > 0\} = \frac{1}{2}$, $E[X_i] = 0$, $E[X_i^2] = 2b^2$, and $\sup_{i \geq 1} E[|X_i|^3] = \rho < \infty$, and let $N \sim \text{Geometric}(p)$ be independent of the X_i 's. Then*

$$d_{BL} \left(\mathcal{L}(p^{\frac{1}{2}} \sum_{i=1}^N X_i), \text{Laplace}(0, b) \right) \leq p^{\frac{1}{2}} \frac{b+2}{b} \left(b\sqrt{2} + \frac{\rho}{6b^2} \right)$$

for all $p \in (0, 1)$.

Theorem 7 gives sufficient conditions for weak convergence in the setting of Theorem 4. Though it requires that the X_i 's have uniformly bounded third absolute moments, the condition of symmetry is weakened to $\mathbb{P}\{X_i < 0\} = \mathbb{P}\{X_i > 0\} = \frac{1}{2}$ and the identical distribution assumption is reduced to the requirement that the X_i 's have common variance. This result is not as general as Theorem 5, but it does provide bounds on the error terms, which is important for practical considerations.

5. CONCLUDING REMARKS

In addition to providing a Stein characterization for the Laplace distribution along with bounds on the error terms from Laplace approximation, and using these to obtain criteria for the convergence of geometric sums to the Laplace distribution with an estimate of the convergence rate, the methodology employed in this paper is of interest in its own right. Firstly, our characterizing equation involves second derivatives and can be obtained by successive applications of the density method. The density method is perhaps the most common means of finding a characterizing equation [3], and it is intriguing that the procedure can be iterated to obtain useful distributional characterizations involving higher derivatives. In the case of the $\text{Laplace}(0, b)$ distribution, the density method suggests the Stein operator $(\mathcal{A}_L g)(x) = g'(x) - \frac{1}{b} \text{sgn}(x)g(x)$, but we were unable to get much mileage out of this characterization while the second-order operator $(\tilde{\mathcal{A}}_L g)(x) = g(x) - g(0) - b^2 g''(x)$

turned out to be quite effective. Perhaps this will also prove to be the case for other distributions. Secondly, our work made use of Goldstein and Reinert's $X - P$ bias transformations [7], which is a very neat theory that has not yet found many applications. It seems likely that as Stein's method techniques are applied to more and more classes of distributions, their results will be useful in constructing corresponding transformations to aid in the calculations. Note also that in this context the present work demonstrates the utility of composing distributional transformations to obtain new ones. Thirdly, we hope that Proposition 1 serves to popularize the use of the bounded Lipschitz metric in the Stein's method community. We were unable to get useful results in the present setting when working directly with the Kolmogorov or Wasserstein distances because of problems with continuity and boundedness issues, but the computations were very nice and easy once we turned to d_{BL} . Moreover, even when one can obtain suitable solutions to the Stein equation using d_W , they are often much simpler if one assumes that the h are bounded as well. Finally, the fact that we were able to parallel the argument from [12] to find a complementary random variable which could be used in Stein's method calculations suggests that this technique may have broader applicability. Just as one can typically find the desired transformation of a sum of a fixed number of random variables by exchanging appropriately chosen summands with their transformations, it appears that one may often be able to deal with sums of random numbers of random variables by introducing a different random variable as the upper index of summation and then exchanging the last summand with its corresponding transformation. It would be interesting to see if this basic strategy works with more general Linnik laws, Mittag-Leffler distributions, and other geo-stable families as well (see [10]). One also hopes that these ideas can be used to find non-geometric random indices which give rise to similar limit theorems.

ACKNOWLEDGMENTS

The authors would like to thank Larry Goldstein for suggesting the use of Stein's method to get convergence rates in the general setting of Theorem 4 and for many helpful comments throughout the preparation of this paper. Thanks also to Alex Rozinov for several illuminating conversations concerning the material in section 4.

REFERENCES

- [1] T.C. BROWN AND M.J. PHILLIPS. Negative binomial approximation with Stein's method. *Methodol. Comput. Appl. Probab.* **1** (1999), no. 4, 407-421.
- [2] L.H.Y. CHEN. Poisson approximation for dependent trials. *Annals of Probability*, **3** (1975), 534– 545.
- [3] L.H.Y. CHEN, L. GOLDSTEIN, AND Q.-M. SHAO. Normal approximation by Stein's method. Probability and its Applications (New York). *Springer, Heidelberg*, 2011. xii+405 pp.
- [4] P. DIACONIS AND S. HOLMES, editors. Stein's method: expository lectures and applications. Papers from the Workshop on Stein's Method held at Stanford University, Stanford, CA, 1998. Institute of Mathematical Statistics Lecture Notes—Monograph Series, 46. *Institute of Mathematical Statistics, Beachwood, OH*, 2004. vi+139 pp.
- [5] R. DURRETT. Probability: theory and examples. Fourth edition. Cambridge Series in Statistical and Probabilistic Mathematics. *Cambridge University Press, Cambridge*, 2010. x+428 pp.
- [6] W. EHM. Binomial approximation to the Poisson binomial distribution. *Statist. Probab. Lett.* **11** (1991), no. 1, 7-16.
- [7] L. GOLDSTEIN AND G. REINERT. Distributional transformations, orthogonal polynomials, and Stein characterizations. *J. Theoret. Probab.* **18** (2005), no. 1, 237-260.
- [8] V. KALASHNIKOV. Geometric sums: bounds for rare events with applications. Risk analysis, reliability, queuing. Mathematics and its Applications, 413. *Kluwer Academic Publishers Group, Dordrecht*, 1997. xviii+265 pp.

- [9] S. KOTZ, T.J. KOZUBOWSKI, AND K. PODGÓRSKI. The Laplace Distribution and Generalizations: A Revisit with Applications to Communications Economics, Engineering, and Finance. *Birkhäuser Boston, Inc., Boston, MA*, 2001. xviii+349 pp.
- [10] T.J. KOZUBOWSKI AND K. PODGÓRSKI. Geometric stable laws through series representations. *Serdica Math. J.* **25** (1999), no. 3, 241-256.
- [11] H.M. LUK. Stein's method for the Gamma distribution and related statistical applications. Thesis (Ph.D.)—University of Southern California. ProQuest LLC, Ann Arbor, MI, 1994. 74 pp.
- [12] E.A. PEKÖZ AND A. RÖLLIN. New rates for exponential approximation and the theorems of Rényi and Yaglom. *Ann. Probab.* **39** (2011), no. 2, 587-608.
- [13] E.A. PEKÖZ, A. RÖLLIN, AND N. ROSS. Degree asymptotics with rates for preferential attachment random graphs. arXiv:1108.5236v1. 2011.
- [14] A. RÉNYI. A characterization of Poisson processes. (Hungarian) *Magyar Tud. Akad. Mat. Kutató Int. Közl.* **1** (1957), 519-527.
- [15] N. ROSS. Fundamentals of Stein's method. *Probab. Surv.* **8** (2011), 210-293.
- [16] C. STEIN. A bound for the error in the normal approximation to the distribution of a sum of dependent random variables. *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability (Univ. California, Berkeley, Calif., 1970/1971), Vol. II: Probability theory*, pp. 583-602. Univ. California Press, Berkeley, Calif., 1972.
- [17] C. STEIN. Approximate computation of expectations. Institute of Mathematical Statistics Lecture Notes—Monograph Series, 7. *Institute of Mathematical Statistics, Hayward, CA*, 1986. iv+164 pp.
- [18] A.A. TODA. Weak limit of the geometric sum of independent but not identically distributed random variables. arXiv:1111.1786v2. 2011.
- [19] A.W. VAN DER VAART AND J.A. WELLNER. Weak convergence and empirical processes. Springer Series in Statistics. *Springer-Verlag, New York*, 1996. 508 pp.